MACHINE LEARNING ASSISTED INTRAOPERATIVE ASSESSMENT OF BRAIN TUMOR MARGINS USING HRMAS NMR SPECTROSCOPY

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

COMPUTER ENGINEERING

By

Doruk Cakmakci

July 2021

Machine Learning Assisted Intraoperative Assessment of Brain Tumor Margins using HRMAS NMR Spectroscopy
By Doruk Cakmakci
July 2021

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

A. Ercument Cicek(Advisor)

Can Alkan

Tolga Can

Approved for the Graduate School of Engineering and Science:

Ezhan Karaşan
Director of the Graduate School

ABSTRACT

MACHINE LEARNING ASSISTED INTRAOPERATIVE ASSESSMENT OF BRAIN TUMOR MARGINS USING HRMAS NMR SPECTROSCOPY

Doruk Cakmakci
M.S. in Computer Engineering
Advisor: A. Ercument Cicek
July 2021

Complete resection of the tumor is important for survival in glioma patients. Even if the gross total resection was achieved, left-over micro-scale tissue in the excision cavity risks recurrence. High Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HRMAS NMR) technique can distinguish healthy and malign tissue efficiently using peak intensities of biomarker metabolites. The method is fast, sensitive and can work with small and unprocessed samples, which makes it a good fit for real-time analysis during surgery. However, only a targeted analysis for the existence of known tumor biomarkers can be made and this requires a technician with chemistry background, and a pathologist with knowledge on tumor metabolism to be present during surgery. Here, we show that we can accurately perform this analysis in real-time and can analyze the full spectrum in an untargeted fashion using machine learning. We work on a new and large HRMAS NMR dataset of glioma and control samples (n = 565), which are also labeled with a quantitative pathology analysis. Our results show that a random forest based approach can distinguish samples with tumor cells and controls accurately and effectively with a median AUC of 85.6% and AUPR of 93.4%. We also show that we can further distinguish benign and malignant samples with a median AUC of 87.1% and AUPR of 96.1%. We analyze the feature (peak) importance for classification to interpret the results of the classifier and validate that known malignancy biomarkers such as creatine and 2-hydroxyglutarate play an important role in distinguishing tumor and normal cells and suggest new biomarker regions.

Keywords: Tumor Margin Detection, HRMAS NMR, Machine Learning, Feature Importance.

ÖZET

BEYİN TÜMÖRÜ SINIRLARININ AMELİYAT SIRASINDA HRMAS NMR SPEKTROSKOPİSİ KULLANILARAK MAKİNE ÖĞRENİMİ DESTEKLİ DEĞERLENDİRİLMESİ

Doruk Cakmakci
Bilgisayar Mühendisliği, Yüksek Lisans
Tez Danışmanı: A. Ercument Cicek
July 2021

Glioma hastalarında tümörün tam rezeksiyonu sağkalım için önemlidir. Gross total rezeksiyon sağlanmış olsa bile, eksizyon kavitesinde kalan mikro ölçekli tümörlü doku nüks riski taşır. Yüksek Çözünürlüklü Sihirli Açı Döndürmeli Nükleer Manyetik Rezonans (HRMAS NMR) tekniği, biyobelirteç metabolit pik uzunluklarını kullanarak sağlıklı ve tümörlü numuneleri başarılı bir şekilde ayırt edebilir. Bu yöntemin küçük ve işlenmemiş numuneler üzerinde hızlı ve hassas bir şekilde çalışması ameliyat sırasında gerçek zamanlı analiz için uygun olmasını sağlar. Ancak, bu metodun ameliyat sırasında kullanımı için kimya bilgisine sahip bir teknisyenin ve tümör metabolizması hakkında bilgi sahibi bir patoloğun ameliyata katılması gereklidir. Ayrıca, bu yöntem ile yapılan analiz bilinen tümör biyobelirteçleri ile kısıtlı ve onların varlığına yöneliktir. Bu çalışmada, bahsedilen analizi HRMAS NMR spektrumu üzerinde makine öğrenimi kullanarak, bilinen tümor biyobelirteç metabolitlerinden bağımsız, gerçek zamanlı ve doğru bir biçimde gerçekleştirebileceğimizi gösterdik. Bu amaç doğrultusunda, glioma ve sağlıklı numuneler içeren, kantitatif bir patoloji analizi ile etiketlenmiş, yeni ve büyük bir HRMAS NMR veri kümesi (n = 565) kullandık. Rassal orman temelli bir yaklaşımın tümorlü ve sağlıklı numuneleri ayırt etmekte başarılı olduğunu (%85.6 medyan AUC ve %93.4 medyan AUPR) gözlemledik. Ayrıca, iyi ve kötü huylu tümörlü numunelerin benzer bir yaklaşım ile ayrılabildiğini (%87.1 medyan AUC ve %96.1 medyan AUPR) gösterdik. Modelimizin sonuçlarını yorumlamak için özellik (pik) önem derecesini analiz ettik. Kreatin ve 2-hidroksiglutarat gibi bilinen malignite biyobelirteçlerinin tümör ve sağlıklı hücreleri ayırt etmede önemli bir rol oynadığını doğruladık ve yeni biyobelirteç bölgeleri önerdik.

Anahtar sözcükler: Tümor Sınır Tespiti, HRMAS NMR, Makine Öğrenimi.

Acknowledgement

First and foremost, I would like to extend my gratitude to my supervisor as well as my mentor Prof. Ercument Cicek not only for demonstrating his interpretation of the working principles of a passionate and successful researcher, but also showing an extraordinary understanding and patience during both good and bad times. I believe we will accomplish many greater things, either as a collective or in a parallel manner, for the betterment of science, humanity and ourselves.

Second, I would like to extend my thanks to Furkan Ozden to the greatest extent possible. Unfortunately, no set of words in English can fully describe the following: my *appreciation for his company* on the road to accomplish many things of great importance, as well as during day-to-day activities; and *admiration* for the person he became and eventually will become. Only a long lasting hug can transmit my feelings towards him in a lossless manner. In this period of our lives, I think we have taken a step towards becoming better versions of ourselves, and towards great accomplishments as well as happiness that it will possibly bring until the bounds of human life or hopefully beyond.

Moreover, I would *not be able* to write this thesis without contributions of my family, including my mother, father and especially my grandmother. Jokes aside, their unconditional support from every point of view, gave me the strength to develop myself and this thesis.

From another important perspective, I would not be able to write *this thesis* if Izzie Jacques Namer MD, PhD and his team in France did not have the vision to collect and store this many tissue specimens while asking each patient for their consent. Many thanks and best wishes!

Finally, I would like to thank jury members, Prof. Can Alkan and Prof. Tolga Can, for accepting our invitation to be a jury member in my defense, and also for their valuable time to read and evaluate this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Gliomas constitute 60% of all primary brain tumors [1]. The maximum resection of the tumor remains the key point in the management of gliomas with a direct influence on the survival of patients [2]. The progress made over the last two decades in surgical techniques including microsurgery by the operating microscope, preoperative functional imaging (e.g., functional MRI, MRI tractography), intraoperative electrical stimulation in awakened craniotomy and intraoperative imaging (surgery guided by real-time imaging using neuronavigation or intraoperative MRI) have largely contributed to significantly increase resected tumor volume while improving morbidity and mortality [3].

Providing feedback on left-over malign tissue during surgery can help surgeons delineate more precisely the limits of a tumor infiltration, especially after a macroscopically complete excision. Several innovative techniques based on optical spectrometry [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16] or mass spectrometry [17, 18, 19, 20, 21, 22, 23, 24, 25] are now proposed to help surgeons to evaluate the margins of resection and possibly to amplify the surgical procedure. Metabolic profiling of a biopsy sample by High Resolution Magic Angle Spinning Nuclear Magnetic Resonance (HRMAS NMR) spectroscopy is a recent novel technique for efficiently distinguishing malign and healthy tissues in excision cavity during surgery. This technique is particularly well-suited for this task due to its ability

to analyze small samples of unprocessed tissue specimens. It has a nondestructive nature and allows other analytical techniques on the same specimen which is important when small amounts of tissue are available [26]. Moreover, the preparation of biopsy samples is fast as it does not require lengthy chemical extraction procedures. Battini *et al.* showed that HRMAS NMR spectroscopy using intact tissue provides solid information in the characterization of pancreatic adenocarcinoma and also on the long-term survival. The information can be obtained in twenty minutes during surgery [27]. A recently released metabolic database on HRMAS NMR signatures of seventy six biomarker metabolites has taken the next step in widening the usage of the technique [28].

One challenge to overcome for this technique to be used in the surgery room is its dependence on human experts with background on chemistry and cancer biology. The raw NMR signal is evaluated by the NMR technician who can report on the existence of certain biomarker metabolites usually with no insight on the tumor metabolism. Evaluation of the raw signal comes with several obstacles. First, the identification of biomarker metabolites might not be possible due to superimposed signature signals of certain metabolites (e.g., creatine and lysine [29]). Second, certain peaks might shift due to experimental conditions (e.g., due to temperature) and then an informed guess on whether that peak belongs to the targeted metabolite must be made. Third, the intra-tumor heterogeneity might result in a convoluted signal and might make it hard for the technician to detect malignant tissue due to unusual relative peak intensities. Moreover, an expert pathologist needs to be present at the time of the surgery to relate the findings of the technicians to the tumor metabolism. Maybe the most restricting factor of this analysis pipeline is the targeted analysis of the raw NMR signal. This means the human expert is limited by the knowledge of certain biomarker metabolites and their corresponding peaks. However, the spectrum contains many uncharacterized regions which might harbor peaks that are capable of distinguishing tumor cells and yet are unknown.

In this study, we propose using machine learning approaches to address the above-mentioned problems and to automate distinguishing healthy tissue from benign/malignant tumor tissue obtained from the excision cavity during tumor

surgery. The algorithm is fast and can work within the time frame of surgery. It directly outputs whether a sample includes tumor cells. Thus, it does not require a technician to analyze the signal. It performs an untargeted analysis of the signal and is able to extract information from uncharacterized regions in the spectrum. The system figure representing the proposed pipeline is shown in Figure 1.1.

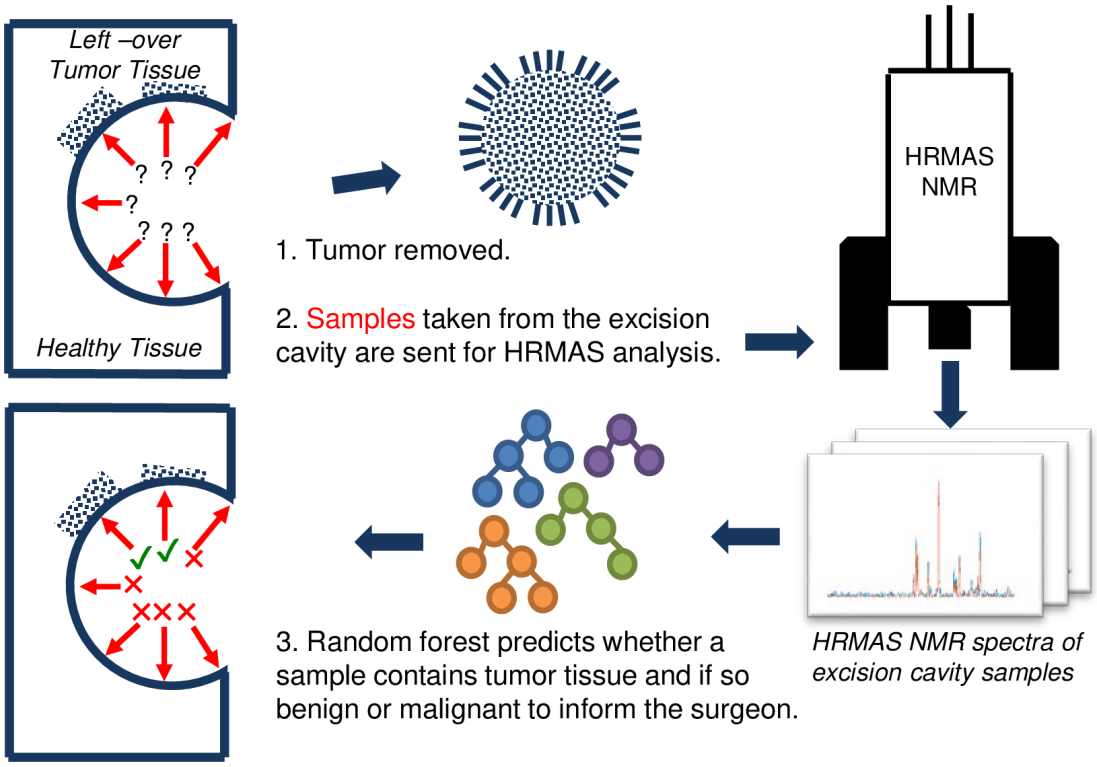

Figure 1.1: The figure shows the pipeline proposed for machine learning assisted tumor margin assessment during brain tumor surgery. After the tumor removal, the surgeon resects samples from the excision cavity. Samples are analyzed via HRMAS NMR technique. Produced spectra are processed via a random forest classifier to label each region in the cavity (malignant/benign tumor vs healthy tissue). The feedback is sent to the surgeon for resecting more tissue for regions labeled positive for tumor tissue.

Here, we utilize a new dataset (n=565) of glioma and control samples analyzed using HRMAS NMR. All samples are also analyzed by a clinical pathologist and labeled whether it is normal, benign or malignant. To the best of our knowledge, this is the largest of its kind with pathological labels. We benchmark various machine learning architectures and show that it is possible to distinguish tumor and control samples with a median AUC of 85.2% and AUPR of 93.4%. We show that

we can also distinguish benign and malignant tumor samples with a median AUC of 87.1% and AUPR of 96.1%. The best performing method is a random forest based approach. This method for the first time performs an untargeted analysis of the spectrum. Moreover, the model is interpretable and informs the user about the ranges in the spectrum that were most informative for the classification, using SHAP values of the features which quantify their importance [30]. We validate that the model focuses on known cancer biomarker metabolites such as creatine and 2-hydroxylglutarate while distinguishing benign and malignant glioma samples. We also observe that branched chain amino acids have been important in the classification. We find evidence in literature that indeed altered branch chain amino acid concentrations are related to glioma metabolism, yet, their status as biomarkers are not well-established. We also find some uncharacterized regions in the spectrum that are informative, which brings up further research questions on establishing an understanding on the compounds in those regions and their relation to tumor metabolism

Chapter 2

Materials and methods

2.1 Ethics Statement

All data used in this study were collected from two sources and approved by the Ethics Committee of Strasbourg (Comité de protection des personnes, Est IV). Tissue specimens were collected, either by a pneumatic system connected between the operating theater of neurosurgery and the NMR room (Hautepierre Hospital - University Hospitals of Strasbourg), or by samples stored in two Tumor Bio-bank, Strasbourg and Colmar (Ethics Committee no. 2003-100, 09.12.2003 and no. 2013-37, 12.11.2013). A written informed consent was obtained from all patients included.

2.2 Dataset

In this subsection, we provide details on the glioma HRMAS NMR dataset and corresponding quantitative pathological analysis to obtain the labels. The dataset is available at https://zenodo.org/record/3951448.

2.2.1 Patient's Cohort and Tissue Sample Collection

The metabolomics-based statistical model was constructed from spectra of 247 primary brain tumor samples from 218 patients, 74 non-tumor brain tissue samples from epilepsy surgery of 54 patients and spectra of 244 samples from excision cavity of patients. The histopathological classification of primary brain tumors is: Pilocytic astrocytoma (AST-I, n=3), astrocytoma grade II (AST-II, n=6), astrocytoma grade III (AST-III, n=5), glioblastoma (GBM, n=123), oligodendroglioma grade II (ODG-II, n=25), oligodendroglioma grade II-III (ODG-II_III, n=7), oligodendroglioma grade III (ODG-III or ODIII, n=41), oligoastrocytoma grade II (OAST-II, n=3), oligoastrocytoma grade III (OAST-III, n=4), oligoastrocytoma grade II-III (OASTII-III, n=1) ganglioglioma grade II (GG-II, n=5), ganglioglioma grade III (GG-III, n=1), dysembryoplastic neuroepithelial tumors (DNET, n=22), gliosarcoma (GS, n=1).

Tissue specimens were collected with minimum ischemic delays after resection (average time $2\,\text{min} \pm 1\,\text{min}$). All tissue samples used in this study had a viable tumor/necrosis ratio and were quantitatively and qualitatively adequate to perform satisfactory NMR HRMAS analysis. In order to wait for this goal, after NMR HRMAS analysis, the inserts were cut, and for half the content of each sample, the percentage of tumor cells in the total sample of cells with regard to the total surface were calculated based on frozen hematoxylin & eosin-stained sections. See Table A.1 for details on collected samples.

2.2.2 HRMAS NMR Data Acquisition

Each brain biopsy sample was prepared at 20°C by introducing a 15- to 18-mg biopsy into a disposable $30\mu L$ KelF insert. To provide a lock frequency for the NMR spectrometer, $10\mu L$ of D_2O was also added to the insert.

All HRMAS NMR spectra were acquired on a Bruker (Karlsruhe, Germany) Avance III 500 spectrometer operating at a proton frequency of 500.13 MHz and

equipped with a 4-mm triple-resonance gradient HRMAS probe (1H, 13C and 31P). The temperature was maintained at 4°C throughout the acquisition time in order to reduce the effects of tissue degradation during the spectrum acquisition. A one-dimensional (1D) proton spectrum using a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence was acquired with a $285\mu s$ inter-pulse delay and a 10-min acquisition time for each tissue sample. The number of loops was set at 328, giving the CPMG pulse train a total length of 93 ms.

2.2.3 HRMAS NMR Data Preprocessing

The free induction decay (FID) signal for each sample had a length of 16,384. The signal is left-shifted by 70 points to remove the Bruker digital filter in the prefix. Obtained raw FID spectrum is then transformed to frequency domain and is phased corrected. The suffix of the signal which contained almost no variance is cropped to obtain the final signal used for analysis, which is of length 8,172. The magnitude of the signal is used for the presented analysis.

2.3 Problem Formulation

In this study, our main task is distinguishing tumor tissue from the normal tissue. The problem is modelled as a binary classification task. For a given HRMAS NMR signal i in the sample set S, the feature vector x_i is a d-dimensional vector: $\vec{x_i} = [x_i^1, x_i^2, \ldots, x_i^d]$ which represents the signal intensity at each ppm. The label for that sample is y_i and $y_i = 1$ if sample contains tumor tissue and $y_i = 0$, otherwise. Then, the model we learn is a function f such that $f(\vec{x_i}) = \hat{y_i}, \forall i \in S$. The second and optional task is to distinguish benign and malignant tumor samples. In this task, a sample j has label $z_j = 1$, if the sample has malignant tumor cells, and $z_j = 0$, if sample contains benign tumor cells. This task is also a binary classification task and we learn a function g such that $g(\vec{x_i}) = \hat{z_i}$. We would like to note that we also considered a multi-class classification task which unites the above mentioned binary classification tasks. However, as we discuss

in Discussion section, we obtained better performance with two separate tasks. Given the first task is of utmost importance and the second is optional, we opted for this approach.

2.4 Learning Algorithms

In this section, we describe the methods employed for the problems formulated above. We benchmark various machine learning algorithms to find the one suitable for the tasks at hand given the size and nature of the 1H HRMAS NMR signal. For all methods, the only input is $\vec{x_i}$ for both tasks ($d = 8,172$). See Experimental Setup section for parameter details of each approach.

First, we run partial least squares discriminant analysis (PLS-DA) as a baseline which is a common method used in metabolomics analysis [31]. As the second algorithm, we used a Random Forest (RF) classifier, which trains many weak classifier trees on sample subsets which are created using bootstrapping [32] and results are aggregating via majority voting. The third algorithm is a support vector machine classifier (SVM). We employed linear and radial basis function (RBF) kernels with a soft margin.

As a baseline neural network architecture, we employ a fully connected multilayer perceptron. Our MLP [33] model takes $\vec{x_i}$ and applies a number of fully connected (FC) layers which makes use of rectified linear unit (ReLU) activation. We use full batch gradient descent for training. At the output layer, we use softmax to assign probabilities to each class (e.g., benign) and focal loss as our loss function to address the class imbalance in our dataset (i.e., smaller number of benign samples) [34].

Convolutional neural networks (CNN) are well-established architectures for learning complex patterns on 2D image data. CNNs have also proven useful for processing 1D data. Some examples are drug chemical structure representation (e.g., SMILES [35, 36]), natural language (i.e., sentences [37]) and EEG

signals [38]. Thus, we conjectured that a CNN is a good candidate for the classification tasks mentioned above. As our final model, we use a 1D CNN-based architecture to process $\vec{x_i}$. The architecture consists of 2 layers of convolutional operations. First, C_1 kernels of size kx1 are passed over the signal with stride s and dilation rate d (no padding). The same set of operations are applied on the output of the first convolutional layer with C_2 kernels. The output is passed through a set of fully connected layers to produce class probabilities using softmax at the output layer. Again, we use full-batch gradient descent as our optimizer and focal loss as our loss function.

Chapter 3

Results

3.1 Experimental Setup

We label the samples in our dataset as aggressive, benign or control using the following method. See Dataset section for details. Per all individuals in the dataset, we have multiple types of samples that originate from (i) the glioma tumor tissue (i.e., glioma), (ii) the healthy brain tissue (i.e., control), and (iii) from the excision cavity (i.e., test). For samples in (i), the aggressive label and the benign label are assigned with respect to the pathological analysis result. For samples in (ii), control label is assigned. For samples in (iii), if the pathology report indicates that tumor cells exist (i.e., positive), then aggressive label is assigned if the tumor of that individual is aggressive and benign label is assigned if the tumor of that individual is benign; otherwise, control label is assigned. In the end, we obtained 179 control, 88 benign and 301 malignant samples. See Table A.1 for details on the labels for collected samples. We generate 2 datasets for the two tasks explained in Problem Formulation section. The first one unites the labels benign and malignant and sets their labels to tumor for task 1. The second one only retains the benign and malignant samples for task 2.

Performance of the proposed models are assessed using a stratified and grouped

8-fold cross validation approach on each dataset. Each dataset is shuffled before the folds are generated. Folds are generated in a stratified manner by sampling from the dataset according to the label distribution of the dataset. That is, each fold has a similar distribution of labels to the whole dataset. There is no sample or patient overlap between the generated folds. That is, an individual's all samples are always in a single fold and the folds are exclusive. In each iteration, first, the test and validation folds are removed. The models are trained on 6 remaining folds and the best performing parameter set is found on the validation fold. Then, each model is trained on 7 folds (training + validation) and is tested on the test fold. This procedure is repeated three times for each task with a random weight initialization of the models. AUC, AUPR distributions are calculated using the performance for each test on each test fold.

For the PLS-DA approach we used 30 components which sets the number of latent variables. For the SVM model we performed a grid search on the soft-margin regularization parameter (i.e., C: 0.01, 0.1, 1, 10, 100) and on the kernel choice (i.e., RBF vs linear). For the RF model, we performed a search on (i) number of estimators: 100, 300, 500, 800, and 1200; (ii) maximum tree depth: 5, 10, 15, 20, 25, and 30; and finally, (iii) minimum number of samples to split a node: 2, 5 10, 15, and 20. We also set the minimum number of samples in a leaf node to 10 to avoid overfitting. For the 4-layered fully connected (baseline) network, the input layer has 8,127 neurons, the second layer has 4,000 neurons, the third layer has 1,000 neurons and the output layer has 2 neurons, which uses softmax to produce probabilities per class in both tasks. ReLU activation is used for all hidden layers. Finally, for the CNN model, we use two convolutional layers such that the number of kernels in both layers are $C_1 = C_2 = 4$. These 1D kernels are of size 16, 32, 64, and 128. We set stride and dilation to 1. After passing through maxpool operations of size 1x4, and ReLU activation, concatenated activation maps are input to fully connected layers which are of size 8,112, 4,000 and 1,000, respectively. Similar to the base neural network model, output layer has 2 nodes with a softmax operation to produce class probabilities and ReLU activation is used for all hidden layers. We trained the networks with a fixed epoch number of 200, which was decided on the validation folds.

All considered machine learning algorithms other than neural networks are implemented in Python language using scikit-learn library. Pytorch framework and Python were used to implement neural networks. All models are trained and tested on a SuperMicro SuperServer 4029GP-TRT with 2 Intel Xeon Gold 6140 Processors (2.3GHz, 24.75M cache), 768GB RAM and 8 NVIDIA GeForce GTX 1080 Ti GPUs (11GB, 352Bit).

3.2 Performance Comparison and the Model of Choice

We compare the performances of the above-mentioned methods using AUC and AUPR metrics. Please see Figure 3.1 for results. For the first task, distinguishing the tumor (glioma) and control cells, all methods perform well and the lowest median AUC achieved is 78.9% and the lowest median AUPR achieved is 87.7% (Figure 3.1, Panel A). We observe that the RF model has the best median AUC value with 85.6% which is $\sim$ 1% improvement over the closest performance by the CNN model. The AUC variance of the RF model is similar to CNN and PLS-DA and smaller than other models. Similarly RF is the best performing model with respect to the AUPR metric with an AUPR of 93.4%. The second best median AUPR is 92.6% and is achieved by CNN model. CNN model has the lowest AUPR variance and RF is the second best. In conclusion, CNN also performs almost as well as RF for this task and is slightly edged by the RF model. RF is a less complex model than CNN and more interpretable compared to CNN. Thus, it is our method of choice for this task.

The second task in our pipeline is optional and is performed when the surgeon also would like to know if the tumor is benign or malignant. Results are shown in Panel B of Figure 3.1. Again, all methods perform well and the lowest median AUC achieved is 80% and the lowest median AUPR achieved is 93.4%. We observe that the RF model has the best median AUC value with 87.1% which is $\sim$ 2% improvement over the closest performance by the CNN model. We also see

that the RF model has the lowest AUC variance among all models. Similarly RF is the best performing model with respect to the AUPR metric with an AUPR of 96.1%. All other methods have a median AUPR of 94%, thus also in this category RF provides a $\sim 2\%$ improvement. The AUPR variance of RF is the lowest and is on par with CNN. Thus, RF is the model of choice for this task as well because of its robustness and high sensitivity and specificity.

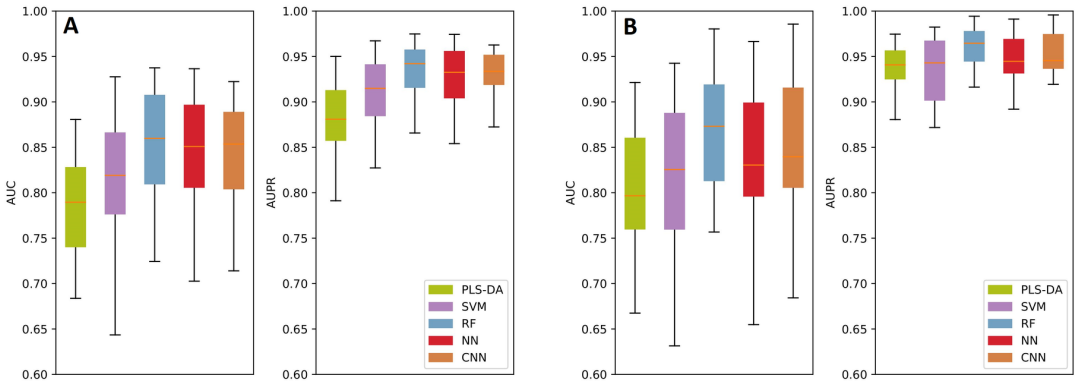

Figure 3.1: The performance comparison of the benchmarked machine learning models in distinguishing control samples and glioma tumor samples (Panel A) and distinguishing malignant and benign glioma tumor samples (Panel B) with respect to the AUC and AUPR metrics. Box plots represent the performance of the models obtained on the test folds, in an 8-fold cross validation setting which is repeated 3 times.

3.3 Interpreting the Model Predictions

We analyze the feature importance of the features (i.e., ppm) that lead to correct classification of the samples in each task with the RF model. For this purpose, we make use of the SHapley Additive exPlanation (SHAP) values for each feature [30, 39]. This approach has its roots in the Shapley values from coalitional game theory. Here, the features are players in a coalition and their values indicate a fair weight that represent their contribution (i.e., success of the classification.)

Here, after running the RF model for both tasks, we compute the SHAP values of each feature (i.e., ppm in the signal) for each task. Here, we map all features

back to the the ppm spectrum (x-axis) and show the corresponding SHAP values (y-axis) for each sample. Each dot on this figure denotes a sample and the color of the sample denotes the value of the corresponding feature. That is, if a sample is purple it means its feature value is high, and if blue, feature value is low. The y-axis (SHAP values) indicates in which direction that feature affects the prediction. That is, for control vs tumor classification task, a positive SHAP value indicates that feature for that sample was important to label it as a tumor sample. On the other hand, a negative SHAP value indicates the feature was important to label it as a control sample. For instance, many purple dots with high SHAP values indicate positive correlation between the tumor and the magnitude of the peak at that ppm. For benign vs malignant classification task, a positive SHAP value indicates malignant label and a negative SHAP value indicates benign label.

We show our results in Figure 3.2. Here, we only annotate the peaks in the SHAP values (most important in either direction) that reach an absolute SHAP value of 0.005. We use the metabolite database provided by Ruhland *et al.* for annotation of peaks. We only list the names of the metabolites which have a group that exactly match with the base of the peak region (i.e., is a subset of the peak region). Note that there are usually many metabolite groups overlap with such regions. To limit the number of candidate metabolites, we use such a stringent criterion. We also annotate the peaks of two well known cancer biomarkers 2-hydroxyglutarate and creatine.

First, we find that 2-hydroxyglutarate has high feature importances in both classification tasks. Isocitrate Dehydrogenase (IDH) is a rate limiting enzyme in the Krebs cycle and plays an important role in the regulation of the energy metabolism. IDH mutations are known to affect tumor metabolism. For instance, mutations of IDH are known to produce high levels of 2-hydroxyglutarate that inhibits glioma stem cell differentiation [40, 41]. So, low levels of 2-hydroxyglutarate indicate malignancy. In line with this information, we observe that when the corresponding peak (feature) values are low (i.e., blue), SHAP values are high which indicates that those samples are predicted to contain tumor and malignant cells, respectively. Similarly, creatine is a well-known biomarker for gliomas. Low creatine levels are observed in IDH mutant gliomas indicating low grade (benign)

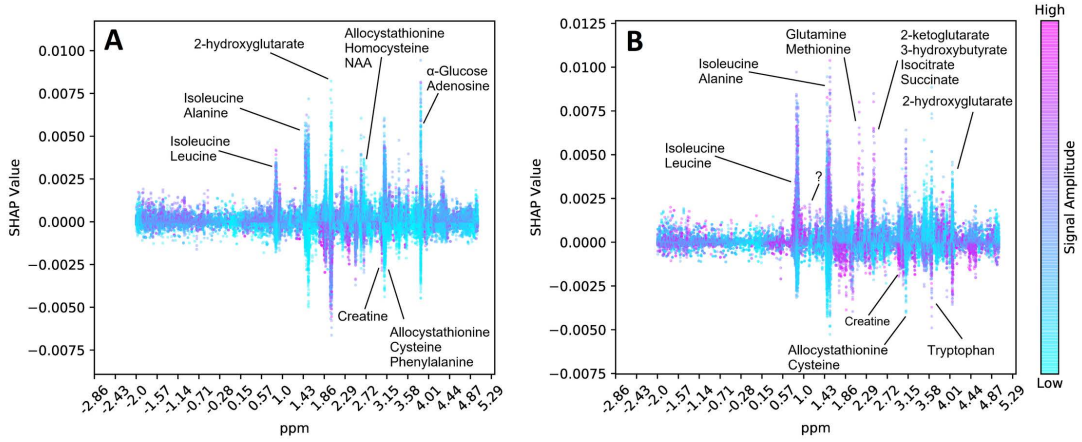

Figure 3.2: The SHAP Values (y-axis) for each ppm in the spectrum (x-axis) is shown for each sample (dots). Dot color purple indicates a high feature value, and blue indicates a low value. A positive SHAP value indicates that feature was important to classify that sample as (i) tumor as opposed to control in Panel A; and as (ii) malignant as opposed to benign in Panel B. Conversely, a negative SHAP value indicates that feature was important to classify that sample as (i) control as opposed to tumor in Panel A; and as (ii) benign as opposed to malignant in Panel B.

tumors [42, 43]. In both tasks, we observe blue peaks with high SHAP values for the ppm range that coincides with creatine groups. This indicates that when creatine levels are low, we predict the sample to be tumor and malignant, respectively. Thus, our model had learnt to focus on regions in the spectrum which are used by technicians today as indicators.

For both tasks, we consistently find that peaks belong to branched chain amino acids isoleucine and leucine are focused by the model. These amino acids are known to have altered concentrations in the presence of IDH mutations, but their status as a biomarker for gliomas are not strongly established. We also observe that various other amino acids are also focused by the model as annotated in Figure 3.2. This suggests possible biomarkers due to the altered amino acid metabolism. Finally, while distinguishing benign and malignant gliomas, we observe that 2-ketoglutarate and Isocitrate are also important factors for successful classification. This is also meaningful as the IDH enzyme catalyzes the reaction

that converts one to other in reversible fashion. IDH mutations affect this process and produce more 2-hydroxyglutarate from 2-ketoglutarate rather then to produce isocitrate [40]. Thus, these are also candidate biomarkers stressed in the prediction of the algorithm.

The interpretation of the results is limited by the 76 metabolites and their ppm signatures provided in [28]. We have performed an analysis to find any SHAP value peaks that are not associated with any metabolite. We obtained top 200 peaks out of $8,172$ and found a relatively short attention peak near 1.00 ppm which indicates malignancy when the concentration is high. This is an uncharacterized region and might suggest a new biomarker. Further research and validation is needed to establish an understanding of the compounds in those regions and their relation to tumor metabolism. Yet, this shows the potential for the untargeted analysis we propose here, as such regions are discarded by an human analyst.

Chapter 4

Discussion

Using a machine learning approach in this application has advantages over a technician commenting on the presence or absence of known biomarker metabolites using the raw signal. Our current catalogue of metabolites in the 1H HRMAS NMR spectrum is limited which means we potentially discard valuable information with this targeted analysis. On the other hand, the RF algorithm we use generates decision tree classifiers, each of which focus on different parts of the spectrum and process features in combinations. Thus, the algorithm performs an untargeted analysis as there is no metabolite identification/quantification. The analysis is also non-linear and multivariate unlike the current approach based on one by one quantification of certain metabolites. Moreover, fluctuations in chemical shift is common in NMR results and a binary guess is needed to conclude whether a peak belongs to a certain metabolite. The RF model can average out such inconsistencies. As seen in Figure 3.2, the focus (i.e., given importance) of the algorithm resembles a peak around certain ppm regions, indicating a smooth adjustment of the weights associated with each ppm, according to the composition in the training cohort. While this untargeted analysis can be performed using many machine learning algorithms, in our detailed benchmark which compares many methods in various settings, we find that RF provides the best results.

Our results provided in Performance Comparison and Model of Choice section

show that our models achieve high AUC and AUPR values indicating that the RF is a viable method to be used in the surgery room. The average test time of the model is negligible (i.e., 0.01 secs.) which makes it possible to use it in real-time. The training phase is performed offline and on average takes 25.2 mins. We interpret the results of the RF model using SHAP values provided for each ppm in the spectrum. We validate that groups of known cancer biomarkers such as creatine and 2-hydroxyglutarate had an important role in the decision made by the model. This is an important feature for this analysis as usually a surgeon would like to know the reasoning behind the decision made by a program. We also indicate several ppm regions which have been important for the classifications. These regions harbor shared groups of several metabolites and further research is needed to validate their ties to glioma metabolism and their status as a glioma biomarker.

We further investigated the value of an untargeted analysis by training another RF model which uses only the top 200 peaks as the sole input (i.e., a feature vector of size 200 instead of 8,172, all other settings are same). For distinguishing healthy and tumor cells, we have seen slight improvements (0.5% and 0.02%) over median AUC and AUPR values compared to the untargetted analysis as some noise features are eliminated. However, using 200 regions in a targeted manual analysis is not feasible. When we go down to using 5 top peaks as input, which is a more manageable size for manual processing, we observe 2.7% and 1.3% decreases in median AUC and AUPR values, respectively. Thus, the untargeted analysis performs better than targeting a small number of peaks as done in the manual analysis and does not require the precious processing time during surgery.

We observe that formulating the problem as a multi-class classification problem and trying to distinguish benign, malignant and control samples does not perform well. The number of benign samples is small and it is hard to distinguish them as their signal resembles the controls. Thus, the median class AUCs we obtained for control and benign samples were down to 60% and 40%, respectively. malignant samples are successfully classified (median AUC =90%) Since, the primary goal is to distinguish tumor and healthy tissue we opted for the presented scheme in this study.

While benchmarking several machine learning algorithms, we observe for both tasks that convolution operation improves the performance of the baseline neural network model slightly and has somewhat lower variance in the performance. Despite being edged by the RF model, we think CNN model can perform well when trained on larger datasets. Our dataset is, to the best of our knowledge, the largest cohort with close to 600 labeled samples. However, CNN uses a deep architecture and requires larger cohorts to learn more complex features. We would like to note that we performed extensive testing on the CNN architecture, which varied the number of layers, number of kernels, activation functions, pooling operations etc. We also experimented with a self-attention mechanism to find regions of interest in the spectrum. The results presented are the best set of results obtained for CNN model. We concluded that the model is too complex to be learnt with this sample size.

Chapter 5

Conclusion

In this thesis, we developed a random forest based machine learning approach to distinguish glioma samples (benign or malignant) from the control samples using the 1H HRMAS NMR signal as the sole input. In our experiments, we show that the approach is efficient, accurate and interpretable. It can work in real-time and thus, can be used as a means of providing feedback to the surgeons on the left-over tumor samples during surgery.

Bibliography

[1] Q. T. Ostrom, H. Gittleman, P. Liao, T. Vecchione-Koval, Y. Wolinsky, C. Kruchko, and J. S. Barnholtz-Sloan, "Cbtrus statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2010–2014," *Neuro-oncology*, vol. 19, no. suppl_5, pp. v1–v88, 2017.

[2] H. J. McCrea, E. D. Bander, R. A. Venn, A. S. Reiner, J. B. Iorgulescu, L. A. Puchi, P. M. Schaefer, G. Cederquist, and J. P. Greenfield, "Sex, age, anatomic location, and extent of resection influence outcomes in children with high-grade glioma," *Neurosurgery*, vol. 77, no. 3, pp. 443–453, 2015.

[3] W. Stummer, M. J. van den Bent, and M. Westphal, "Cytoreductive surgery of glioblastoma as the key to successful adjuvant therapies: new arguments in an old discussion," *Acta neurochirurgica*, vol. 153, no. 6, pp. 1211–1218, 2011.

[4] W. Stummer, U. Pichlmeier, T. Meinel, O. D. Wiestler, F. Zanella, H.-J. Reulen, A.-G. S. Group, *et al.*, "Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase iii trial," *The lancet oncology*, vol. 7, no. 5, pp. 392–401, 2006.

[5] A. Tsugu, H. Ishizaka, Y. Mizokami, T. Osada, T. Baba, M. Yoshiyama, J. Nishiyama, and M. Matsumae, "Impact of the combination of 5-aminolevulinic acid–induced fluorescence with intraoperative magnetic resonance imaging–guided surgery for glioma," *World neurosurgery*, vol. 76, no. 1-2, pp. 120–127, 2011.

[6] M. J. Colditz and R. L. Jeffree, "Aminolevulinic acid (ala)–protoporphyrin ix fluorescence guided tumour resection. part 1: Clinical, radiological and pathological studies," *Journal of Clinical Neuroscience*, vol. 19, no. 11, pp. 1471–1474, 2012.

[7] B. Montcel, L. Mahieu-Williame, X. Armoiry, D. Meyronet, and J. Guyotat, "Two-peaked 5-ala-induced ppix fluorescence emission spectrum distinguishes glioblastomas from low grade gliomas and infiltrative component of glioblastomas," *Biomedical optics express*, vol. 4, no. 4, pp. 548–558, 2013.

[8] Y. Li, R. Rey-Dios, D. W. Roberts, P. A. Valdés, and A. A. Cohen-Gadol, "Intraoperative fluorescence-guided resection of high-grade gliomas: a comparison of the present techniques and evolution of future strategies," *World neurosurgery*, vol. 82, no. 1-2, pp. 175–185, 2014.

[9] F.-K. Lu, D. Calligaris, O. I. Olubiyi, I. Norton, W. Yang, S. Santagata, X. S. Xie, A. J. Golby, and N. Y. Agar, "Label-free neurosurgical pathology with stimulated raman imaging," *Cancer research*, vol. 76, no. 12, pp. 3451–3462, 2016.

[10] M. Jermyn, J. Mercier, K. Aubertin, J. Desroches, K. Urmey, J. Karamchandiani, E. Marple, M.-C. Guiot, F. Leblond, and K. Petrecca, "Highly accurate detection of cancer in situ with intraoperative, label-free, multimodal optical spectroscopy," *Cancer research*, vol. 77, no. 14, pp. 3942–3950, 2017.

[11] M. Jermyn, J. Desroches, J. Mercier, K. St-Arnaud, M.-C. Guiot, F. Leblond, and K. Petrecca, "Raman spectroscopy detects distant invasive brain cancer cells centimeters beyond mri capability in humans," *Biomedical optics express*, vol. 7, no. 12, pp. 5129–5137, 2016.

[12] D. T. M. Chan, H. Y.-P. Sonia, and W. S. Poon, "5-aminolevulinic acid fluorescence guided resection of malignant glioma: Hong kong experience," *Asian journal of surgery*, vol. 41, no. 5, pp. 467–472, 2018.

[13] D. A. Orringer, B. Pandian, Y. S. Niknafs, T. C. Hollon, J. Boyle, S. Lewis, M. Garrard, S. L. Hervey-Jumper, H. J. Garton, C. O. Maher, *et al.*, "Rapid

intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated raman scattering microscopy," *Nature biomedical engineering*, vol. 1, no. 2, p. 0027, 2017.

[14] F. Poulon, H. Mehidine, M. Juchaux, P. Varlet, B. Devaux, J. Pallud, and D. A. Haidar, "Optical properties, spectral, and lifetime measurements of central nervous system tumors in humans," *Scientific reports*, vol. 7, no. 1, pp. 1–8, 2017.

[15] T. C. Hollon, S. Lewis, B. Pandian, Y. S. Niknafs, M. R. Garrard, H. Garton, C. O. Maher, K. McFadden, M. Snuderl, A. P. Lieberman, *et al.*, "Rapid intraoperative diagnosis of pediatric brain tumors using stimulated raman histology," *Cancer research*, vol. 78, no. 1, pp. 278–289, 2018.

[16] Z. Xue, L. Kong, C.-c. Pan, Z. Wu, J.-t. Zhang, and L.-w. Zhang, "Fluorescein-guided surgery for pediatric brainstem gliomas: Preliminary study and technical notes," *Journal of Neurological Surgery Part B: Skull Base*, vol. 79, no. S 04, pp. S340–S346, 2018.

[17] M. V. Brown, J. E. McDunn, P. R. Gunst, E. M. Smith, M. V. Milburn, D. A. Troyer, and K. A. Lawton, "Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies," *Genome medicine*, vol. 4, no. 4, p. 33, 2012.

[18] K.-C. Schafer, J. Balog, T. Szaniszló, D. Szalay, G. Mezey, J. Dénes, L. Bognár, M. Oertel, and Z. Takáts, "Real time analysis of brain tissue by direct combination of ultrasonic surgical aspiration and sonic spray mass spectrometry," *Analytical Chemistry*, vol. 83, no. 20, pp. 7729–7735, 2011.

[19] L. S. Eberlin, I. Norton, D. Orringer, I. F. Dunn, X. Liu, J. L. Ide, A. K. Jarmusch, K. L. Ligon, F. A. Jolesz, A. J. Golby, *et al.*, "Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors," *Proceedings of the National Academy of Sciences*, vol. 110, no. 5, pp. 1611–1616, 2013.

[20] V. Pirro, C. M. Alfaro, A. K. Jarmusch, E. M. Hattab, A. A. Cohen-Gadol, and R. G. Cooks, "Intraoperative assessment of tumor margins during glioma

resection by desorption electrospray ionization-mass spectrometry," *Proceedings of the National Academy of Sciences*, vol. 114, no. 26, pp. 6700–6705, 2017.

[21] S. Santagata, L. S. Eberlin, I. Norton, D. Calligaris, D. R. Feldman, J. L. Ide, X. Liu, J. S. Wiley, M. L. Vestal, S. H. Ramkissoon, *et al.*, "Intraoperative mass spectrometry mapping of an onco-metabolite to guide brain tumor surgery," *Proceedings of the National Academy of Sciences*, vol. 111, no. 30, pp. 11121–11126, 2014.

[22] D. Calligaris, D. R. Feldman, I. Norton, P. K. Brastianos, I. F. Dunn, S. Santagata, and N. Y. Agar, "Molecular typing of meningiomas by desorption electrospray ionization mass spectrometry imaging for surgical decision-making," *International journal of mass spectrometry*, vol. 377, pp. 690–698, 2015.

[23] D. Calligaris, D. R. Feldman, I. Norton, O. Olubiyi, A. N. Changelian, R. Machaidze, M. L. Vestal, E. R. Laws, I. F. Dunn, S. Santagata, *et al.*, "Maldi mass spectrometry imaging analysis of pituitary adenomas for near-real-time tumor delineation," *Proceedings of the National Academy of Sciences*, vol. 112, no. 32, pp. 9978–9983, 2015.

[24] A. K. Jarmusch, V. Pirro, Z. Baird, E. M. Hattab, A. A. Cohen-Gadol, and R. G. Cooks, "Lipid and metabolite profiles of human brain tumors by desorption electrospray ionization-ms," *Proceedings of the National Academy of Sciences*, vol. 113, no. 6, pp. 1486–1491, 2016.

[25] B. Fatou, P. Saudemont, E. Leblanc, D. Vinatier, V. Mesdag, M. Wisztorski, C. Focsa, M. Salzet, M. Ziskind, and I. Fournier, "In vivo real-time mass spectrometry for guided surgery application," *Scientific reports*, vol. 6, no. 1, pp. 1–14, 2016.

[26] M. Gogiashvili, J. Nowacki, R. Hergenröder, J. G. Hengstler, J. Lambert, and K. Edlund, "Hr-mas nmr based quantitative metabolomics in breast cancer," *Metabolites*, vol. 9, no. 2, p. 19, 2019.

[27] S. Battini, F. Faitot, A. Imperiale, A. Cicek, C. Heimburger, G. Averous, P. Bachellier, and I. Namer, "Metabolomics approaches in pancreatic adenocarcinoma: tumor metabolism profiling predicts clinical outcome of patients," *BMC medicine*, vol. 15, no. 1, p. 56, 2017.

[28] E. Ruhland, C. Bund, H. Outilaft, M. Piotto, and I.-J. Namer, "A metabolic database for biomedical studies of biopsy specimens by high-resolution magic angle spinning nuclear mr: a qualitative and quantitative tool," *Magnetic resonance in medicine*, vol. 82, no. 1, pp. 62–83, 2019.

[29] E. O. Karakaslar, B. Coskun, H. Outilaft, I. J. Namer, and E. Cicek, "Predicting carbon spectrum in heteronuclear single quantum coherence spectroscopy for online feedback during surgery," *IEEE/ACM transactions on computational biology and bioinformatics*, 2019.

[30] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, pp. 4765–4774, 2017.

[31] B. Worley and R. Powers, "Multivariate analysis in metabolomics," *Current Metabolomics*, vol. 1, no. 1, pp. 92–107, 2013.

[32] A. Liaw, M. Wiener, *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[33] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 683–697, 1992.

[34] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[35] M. Hirohara, Y. Saito, Y. Koda, K. Sato, and Y. Sakakibara, "Convolutional neural network based on smiles representation of compounds for detecting chemical motif," *BMC bioinformatics*, vol. 19, no. 19, p. 526, 2018.

[36] O. C. Uner, R. G. Cinbis, O. Tastan, and A. E. Cicek, "Deepside: A deep learning framework for drug side effect prediction," *bioRxiv*, p. 843029, 2019.

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[38] Y. Sun, F. P.-W. Lo, and B. Lo, "Eeg-based user identification system using 1d-convolutional long short-term memory neural networks," *Expert Systems with Applications*, vol. 125, pp. 259–267, 2019.

[39] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 2522–5839, 2020.

[40] J. Huang, J. Yu, L. Tu, N. Huang, H. Li, and Y. Luo, "Isocitrate dehydrogenase mutations in glioma: From basic discovery to therapeutics development," *Frontiers in oncology*, vol. 9, p. 506, 2019.

[41] M. S. Waitkus, B. H. Diplas, and H. Yan, "Isocitrate dehydrogenase mutations in gliomas," *Neuro-oncology*, vol. 18, no. 1, pp. 16–26, 2015.

[42] H. Zhao, A. B. Heimberger, Z. Lu, X. Wu, T. R. Hodges, R. Song, and J. Shen, "Metabolomics profiling in plasma samples from glioma patients correlates with tumor phenotypes," *Oncotarget*, vol. 7, no. 15, p. 20486, 2016.

[43] H. Yerli, A. M. Agildere, Ö. Özen, E. Geyik, B. Atalay, and A. H. Elhan, "Evaluation of cerebral glioma grade by using normal side creatine as an internal reference in multi-voxel 1h-mr spectroscopy," *Diagnostic and interventional radiology*, vol. 13, no. 1, p. 3, 2007.

Appendix A

Supplementary Tables

Title Explanations

Identifier: The unique identifier for tissue specimens.
Group: The grouping in order to distinguish primary brain tumors (GLIOMA), excision cavity samples (TEST) and healthy samples.
Pathology: Pathology of the sample. The field includes the tumor subtype and grade for tumor samples and availability of tumor tissue in the specimen for excision cavity samples (see Section 3.1).
Pathologic Classification: Tissue specimen labeling used as ground truth. Possible values are Aggressive-GLIOMA, Benign-GLIOMA, Control. Pathologic classification of excision cavity samples are determined with respect to the primary tumor sample extracted from the same patient.

Table A.1: This table contains the meta-data about samples in the dataset. Specifically (i) information about sample identifiers, groups and pathologic classification; and (ii) identifiers of samples in each dataset fold are provided.

Identifier	Group	Pathology	Pathologic classification
0013MIC-AST	GLIOMA	AST-III	Aggressive-GLIOMA
0015A-GF-AST	GLIOMA	AST-III	Aggressive-GLIOMA
0014A-SJ-AST	GLIOMA	AST-II	Benign-GLIOMA
0015PEN-AST	GLIOMA	AST-I	Benign-GLIOMA
0001HVA-AST	GLIOMA	AST-I	Benign-GLIOMA
0002GFR-AST	GLIOMA	OAST-III	Aggressive-GLIOMA
0003RAR-AST	GLIOMA	AST-II	Benign-GLIOMA
0004PCL-AST	GLIOMA	AST-I	Benign-GLIOMA
0022SRE-MLF	CONTROL	CTX	Control
0023DAS-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0023DAS-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0024DAR-SCL1	CONTROL	CTX	Control
0024DAR-SCL2	CONTROL	CTX	Control
0025FER-GGG1	GLIOMA	GG-II	Benign-GLIOMA
0025FER-GGG2	GLIOMA	GG-II	Benign-GLIOMA
0026WIO-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0026WIO-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0028KLA-EPI1	CONTROL	CTX	Control
0028KLA-EPI2	CONTROL	CTX	Control
0028KLA-EPI3	CONTROL	CTX	Control
0029KUH-SCL1	CONTROL	CTX	Control
0029KUH-SCL2	CONTROL	CTX	Control
0029KUH-SCL3	CONTROL	CTX	Control
0031YEH-SCL10	CONTROL	CTX	Control
0033WER-SCL1	CONTROL	CTX	Control
0033WER-SCL3	CONTROL	CTX	Control
0034SAN-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0035NLI-SCL2	CONTROL	CTX	Control

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0036KEA-SCL1	CONTROL	CTX	Control
0036KEA-SCL3	CONTROL	CTX	Control
0038DEG-EPI	CONTROL	CTX	Control
0039IAM-EPI1	CONTROL	CTX	Control
0039IAM-EPI2	CONTROL	CTX	Control
0040FAT-SCL2	CONTROL	CTX	Control
0043LOR-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0045KGA-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0046WIO-DNET	GLIOMA	DNET	Benign-GLIOMA
0047MGO-DNET	GLIOMA	DNET	Benign-GLIOMA
0049ZIO-SCL2	CONTROL	CTX	Control
0054DEL-SCL2	CONTROL	CTX	Control
0055PAE-GGG1	GLIOMA	GG-II	Benign-GLIOMA
0055PAE-GGG2	GLIOMA	GG-II	Benign-GLIOMA
0055PAE-GGG3	GLIOMA	GG-II	Benign-GLIOMA
0059AER-MLF2	CONTROL	CTX	Control
0060BRI-EPI2	CONTROL	CTX	Control
0061BIS-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0062FLI-EPI3	CONTROL	CTX	Control
0063PAT-SCL1	CONTROL	CTX	Control
0063PAT-SCL2	CONTROL	CTX	Control
0065HIL-EPI2	CONTROL	CTX	Control
0068KHI-SCL2	CONTROL	CTX	Control
0072PRA-MLF2	CONTROL	CTX	Control
0066GAY-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0044JER-EPI2	CONTROL	CTX	Control
0040FAT-SCL1	CONTROL	CTX	Control
0069BAT-EPI3	CONTROL	CTX	Control
0069BAT-EPI10	CONTROL	CTX	Control
0031YEH-SCL20	CONTROL	CTX	Control
0031YEH-SCL30	CONTROL	CTX	Control

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0033WER-SCL2	CONTROL	CTX	Control
0034SAN-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0035NLI-SCL1	CONTROL	CTX	Control
0036KEA-SCL2	CONTROL	CTX	Control
0066GAY-DNET-3	GLIOMA	DNET	Benign-GLIOMA
0044JER-EPI1	CONTROL	CTX	Control
0068KHI-SCL1	CONTROL	CTX	Control
0061BIS-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0049ZIO-SCL1	CONTROL	CTX	Control
0059AER-MLF1	CONTROL	CTX	Control
0060BRI-EPI1	CONTROL	CTX	Control
0054DEL-SCL1	CONTROL	CTX	Control
0062FLI-EPI1	CONTROL	CTX	Control
0065HIL-EPI1	CONTROL	CTX	Control
0066GAY-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0069BAT-EPI30	CONTROL	CTX	Control
0043LOR-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0048NLI-SCL	CONTROL	CTX	Control
0072PRA-MLF1	CONTROL	CTX	Control
0069BAT-EPI2	CONTROL	CTX	Control
0017HAR-GBM-0	GLIOMA	GBM	Aggressive-GLIOMA
0064HR0-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0025SEN-GBM-0	GLIOMA	GBM	Aggressive-GLIOMA
0054GEA-GBM-0	GLIOMA	GBM	Aggressive-GLIOMA
0065SOU-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0068DC1-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0069BRE-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0070SVO-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0071B00-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0072LDR-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0073FON-GBM	GLIOMA	GBM	Aggressive-GLIOMA

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0074KNT-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0075GAT-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0076CAT-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0077RVE-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0078TLA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0079WEN-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0080BID-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0081BAN-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0082FON-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0083SAB-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0084BHR-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0085MRA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0086NUS-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0087EEL-GBM-0	GLIOMA	GBM	Aggressive-GLIOMA
0088SRA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0004A-MC-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0009A-AC-GBM-bis	GLIOMA	GBM	Aggressive-GLIOMA
0012A-HA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0013A-BJ-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0016A-GF-GBM-bis	GLIOMA	GBM	Aggressive-GLIOMA
0017A-GC-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0018A-AJ-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0019A-GF-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0020A-GF-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0013LJE-GBM-00	GLIOMA	GBM	Aggressive-GLIOMA
0007BMA-GBM-0	GLIOMA	GBM	Aggressive-GLIOMA
0012FMA-GBM-0	GLIOMA	GBM	Aggressive-GLIOMA
0001CNA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0002DCH-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0003MGE-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0004SRO-GBM	GLIOMA	GBM	Aggressive-GLIOMA

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0005SAN-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0006KMO-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0008VMA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0009ZMI-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0010BJO-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0011AHE-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0014BPI-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0015LDA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0016MCH-GBM-0	GLIOMA	GBM	Aggressive-GLIOMA
0017SYV-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0018DPI-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0019ZNA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0020MCL-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0021GRI-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0022RRI-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0023SJA-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0024EJE-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0025DRE-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0005A-BA-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0007A-WD-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0008A-OD-ODII_III	GLIOMA	ODG-II_III	Aggressive-GLIOMA
0002ODU-ODII_III	GLIOMA	ODG-II_III	Aggressive-GLIOMA
0001A-SC-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0002A-MC-ODIII-bis	GLIOMA	ODG-III	Aggressive-GLIOMA
0006A-GJ-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0011A-WS-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0021A-BL-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0010A-MD-OD	GLIOMA	OAST-III	Aggressive-GLIOMA
0055MON-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0056BIM-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0057FIM-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0058OER-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0059BLI-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0061HHR-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0062FD1-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0063DAN-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0064MUG-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0065LIN-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0066MIL-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0067BER-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0068BAU-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0069HIR-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0070NAT-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0071NAT-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0073AEL-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0075MRA-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0077DNN-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0001AMA-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0003LHE-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0004SAN-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0005KNA-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0006MMI-ODIII	GLIOMA	ODIII_GBM	Aggressive-GLIOMA
0007BFA-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0008SCH-ODII_III	GLIOMA	ODG-II_III	Aggressive-GLIOMA
0009AAR-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0012SDI-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0013BAL-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0014BLI-ODII	GLIOMA	OAST-II	Benign-GLIOMA
0015KAN-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0016THO-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0017SJA-ODII_III	GLIOMA	ODG-II_III	Aggressive-GLIOMA
0018LPI-ODII	GLIOMA	ODG-II	Benign-GLIOMA

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0019HCA-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0020LMA-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0021VTH-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0022BMA-ODII	GLIOMA	ODG-II	Benign-GLIOMA
0023LMA-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0024NTH-ODII_III	GLIOMA	ODG-II_III	Aggressive-GLIOMA
0025GYV-ODII	GLIOMA	OAST-II	Benign-GLIOMA
0026GAL-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0027WBE-ODIII	GLIOMA	OAST-III	Aggressive-GLIOMA
0004ALF-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0005ADB-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0006ADB-GBM-2	TEST	neg	-
0007ADB-GBM-3	TEST	neg	-
0008ADB-GBM-4	TEST	neg	-
0009ADB-GBM-5	TEST	neg	-
0010ADB-GBM-6	TEST	neg	-
0016ARJ-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0017ARJ-GBM-2-bis	TEST	neg	-
0018ARJ-GBM-3	TEST	neg	-
0054APE-GBM-4	GLIOMA	GBM	Aggressive-GLIOMA
0051APE-GBM-1	TEST	neg	-
0052APE-GBM-2	TEST	pos	-
0053APE-GBM-3	TEST	neg	-
0055APE-GBM-5	TEST	neg	-
0056APE-GBM-6	TEST	neg	-
0063ADP-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0062ADP-GBM-1	TEST	pos	-
0064ADP-GBM-3	TEST	pos	-
0065ADP-GBM-4	TEST	pos	-
0007BDE-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0008BDE-GBM-2	TEST	pos	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0009BDE-GBM-3	TEST	pos	-
0010BDE-GBM-4	TEST	pos	-
0011BDE-GBM-5	TEST	pos	-
0012BDE-GBM-6	TEST	pos	-
0013BRC-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0014BRC-GBM-2	TEST	pos	-
0015BRC-GBM-3	TEST	pos	-
0016BRC-GBM-4	TEST	pos	-
0017BRC-GBM-5	TEST	pos	-
0018BCY-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0019BCY-GBM-2	TEST	pos	-
0020BCY-GBM-3-bis	TEST	pos	-
0021BCY-GBM-4-bis	TEST	pos	-
0022BCY-GBM-5-0	TEST	pos	-
0025BBD-GBM-4	GLIOMA	GBM	Aggressive-GLIOMA
0023BBD-GBM-1	TEST	neg	-
0024BBD-GBM-3	TEST	pos	-
0027BFM-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0028BFM-GBM-2	TEST	neg	-
0029BFM-GBM-3	TEST	neg	-
0030BFM-GBM-4-bis	TEST	neg	-
0031BCS-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0032BCS-GBM-2	TEST	pos	-
0033BCS-GBM-3	TEST	pos	-
0034BCS-GBM-4	TEST	pos	-
0035BCS-GBM-5	TEST	pos	-
0036BCS-GBM-6	TEST	pos	-
0038BLO-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0037BLO-GBM-1-bis	TEST	neg	-
0039BLO-GBM-3	TEST	pos	-
0040BLO-GBM-4	TEST	pos	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0042BPD-GBM-2-0	GLIOMA	GBM	Aggressive-GLIOMA
0041BPD-GBM-1	TEST	neg	-
0042BPD-GBM-2	TEST	pos	-
0043BPD-GBM-3	TEST	pos	-
0044BPD-GBM-4	TEST	neg	-
0010PT4-ODII-1	GLIOMA	ODG-II	Benign-GLIOMA
0011PT4-ODII-2	TEST	pos	-
0012PT4-ODII-3	TEST	neg	-
0002ASE-ODIII-2	GLIOMA	ODG-III	Aggressive-GLIOMA
0001ASE-ODIII-1	TEST	pos	-
0003ASE-ODIII-3	TEST	pos	-
0020ASH-ODII-2	TEST	neg	-
0021ASH-ODII-3	TEST	neg	-
0022AVD-ODIII-1	GLIOMA	ODG-III	Aggressive-GLIOMA
0023AVD-ODIII-2	TEST	pos	-
0024AVD-ODIII-3	TEST	neg	-
0025AVD-ODIII-4	TEST	pos	-
0026AVD-ODIII-5	TEST	pos	-
0030ASG-ODIII-1	GLIOMA	ODG-III	Aggressive-GLIOMA
0031ASG-ODIII-2	TEST	neg	-
0032ASG-ODIII-3	TEST	pos	-
0033ASG-ODIII-4	TEST	pos	-
0034ASG-ODIII-5	TEST	pos	-
0035ARJ-ODIII-1-0	GLIOMA	ODG-III	Aggressive-GLIOMA
0036ARJ-ODIII-2-0	TEST	pos	-
0037ARJ-ODIII-3-0	TEST	neg	-
0038ARJ-ODIII-4	TEST	pos	-
0039ARJ-ODIII-5	TEST	pos	-
0045ALC-ODII-1	GLIOMA	ODG-II	Benign-GLIOMA
0046ALC-ODII-2	TEST	pos	-
0047ALC-ODII-3-0	TEST	neg	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0048ALC-ODII-4	TEST	neg	-
0049ABC-ODII-1	TEST	neg	-
0050ABC-ODII-2	TEST	neg	-
0001BUC-ODII-1	GLIOMA	ODG-II	Benign-GLIOMA
0002BUC-ODII-2	TEST	neg	-
0003BUC-ODII-3	TEST	neg	-
0004BUC-ODII-4	TEST	neg	-
0005BUC-ODII-5	TEST	neg	-
0006BUC-ODII-6	TEST	neg	-
0027AJL-RN-1	TEST	neg	-
0028AJL-RN-2	TEST	neg	-
0029AJL-RN-3	TEST	neg	-
0044AZF-GGAB-5	GLIOMA	GG-III	Aggressive-GLIOMA
0040AZF-GGAB-1-0	TEST	pos	-
0041AZF-GGAB-2-0	TEST	neg	-
0042AZF-GGAB-3	TEST	neg	-
0043AZF-GGAB-4	TEST	pos	-
0057AGP-GS-1	GLIOMA	GS	Aggressive-GLIOMA
0058AGP-GS-2	TEST	pos	-
0059AGP-GS-3	TEST	pos	-
0060AGP-GS-4	TEST	pos	-
0061AGP-GS-5	TEST	neg	-
0013AWR-GS-2	GLIOMA	GS	Aggressive-GLIOMA
0011AWR-GS-5	TEST	neg	-
0012AWR-GS-1	TEST	neg	-
0014AWR-GS-3	TEST	neg	-
0008AGN-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0009AGN-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0010AHP-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0011AHP-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0002ADA-EPI-2	CONTROL	CTX	Control

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0003ADA-EPI-3	CONTROL	CTX	Control
0004ADA-EPI-4	CONTROL	SB	Control
0005ANS-EPI-1	CONTROL	CTX	Control
0006AOJ-SCL-1	CONTROL	SB	Control
0007AOJ-SCL-2	CONTROL	CTX	Control
0012ARJ-SCL-1	CONTROL	SB	Control
0013ARJ-SCL-2	CONTROL	CTX	Control
0015ARS-EPI-2	CONTROL	SB	Control
0016ABE-SCL-1	CONTROL	CTX	Control
0017ABE-SCL-2	CONTROL	CTX	Control
0018APH-EPI-1	CONTROL	SB	Control
0019APH-EPI-2	CONTROL	CTX	Control
0020ABF-SCL-1	CONTROL	CTX	Control
0022ABG-MLF-1	CONTROL	CTX	Control
0023ABG-MLF-2	CONTROL	CTX	Control
0024AVA-EPI-1	CONTROL	SB	Control
0025AVA-EPI-2	CONTROL	CTX	Control
0028AGM-SCL-1	CONTROL	CTX	Control
0034ABC-SCL-1-0	CONTROL	CTX	Control
0027ASP-SCL-2-0	CONTROL	SB	Control
0026ASP-EPI-1-00	CONTROL	CTX	Control
0048ABG-DNET-1	GLIOMA	DNET	Benign-GLIOMA
0049ABG-DNET-2	GLIOMA	DNET	Benign-GLIOMA
0066ASC-ODII-1	GLIOMA	ODG-II	Benign-GLIOMA
0067ASC-ODII-2	TEST	neg	-
0069ASC-ODII-4	TEST	neg	-
0068ASC-ODII-3-0	TEST	neg	-
0070AKS-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0075AOM-ODII-5	GLIOMA	ODG-II	Benign-GLIOMA
0073AOM-ODII-3	TEST	pos	-
0071AOM-ODII-1	TEST	neg	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0072AOM-ODII-2	TEST	neg	-
0077ARA-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0076ARA-GBM-1	TEST	neg	-
0078ARA-GBM-3	TEST	pos	-
0082ABY-GBM-4-0	GLIOMA	GBM	Aggressive-GLIOMA
0079ABY-GBM-1	TEST	neg	-
0081ABY-GBM-3	TEST	neg	-
0080ABY-GBM-2-0	TEST	neg	-
0083ASN-GBM-1	TEST	neg	-
0087ASN-GBM-5	TEST	neg	-
0088ASN-GBM-6	TEST	neg	-
0084ASN-GBM-2-0	TEST	neg	-
0089APH-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0091APH-GBM-3	TEST	neg	-
0092APH-GBM-4	TEST	neg	-
0090APH-GBM-2-0	TEST	pos	-
0093AVT-AST-1	GLIOMA	AST-II	Benign-GLIOMA
0096AVT-AST-4	TEST	pos	-
0094AVT-AST-2	TEST	pos	-
0095AVT-AST-3-0	TEST	pos	-
0097ALD-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0104AWJ-GBM-3	TEST	neg	-
0105AWJ-GBM-4	TEST	neg	-
0106AWJ-GBM-5	TEST	neg	-
0112ARM-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0113ARM-GBM-2	TEST	neg	-
0114ARM-GBM-3	TEST	pos	-
0115ARM-GBM-4	TEST	pos	-
0116ARM-GBM-5	TEST	neg	-
0124ADJ-GBM-1-0	TEST	neg	-
0125ADJ-GBM-2	TEST	neg	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0126ADJ-GBM-3	TEST	neg	-
0127ADJ-GBM-4	TEST	neg	-
0128AVA-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0130AVA-GBM-3	TEST	neg	-
0131AVA-GBM-4	TEST	neg	-
0107AHR-GBM-1-0	GLIOMA	GBM	Aggressive-GLIOMA
0134AMJ-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0138AKS-ODIII-4	GLIOMA	ODG-III	Aggressive-GLIOMA
0136AKS-ODIII-2	TEST	neg	-
0137AKS-ODIII-3	TEST	neg	-
0139AAR-AST-1	GLIOMA	AST-II	Benign-GLIOMA
0140AAR-AST-2	TEST	pos	-
0141AAR-AST-3	TEST	pos	-
0142AAR-AST-4	TEST	pos	-
0143ABM-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0144ABM-GBM-2	TEST	pos	-
0145ABM-GBM-3	TEST	neg	-
0146ABM-GBM-4	TEST	pos	-
0148AFG-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0147AFG-GBM-1	TEST	neg	-
0149AFG-GBM-3	TEST	neg	-
0150AFG-GBM-4	TEST	neg	-
0151AFG-GBM-5	TEST	neg	-
0152AFG-GBM-6	TEST	neg	-
0153AWA-AST-1	GLIOMA	AST-III	Aggressive-GLIOMA
0154AWA-AST-2	TEST	pos	-
0157ASS-GBM-3	GLIOMA	GBM	Aggressive-GLIOMA
0155ASS-GBM-1	TEST	pos	-
0156ASS-GBM-2	TEST	neg	-
0158ASS-GBM-4	TEST	neg	-
0161ADR-AST-1	GLIOMA	AST-III	Aggressive-GLIOMA

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0166AVA-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0167AVA-GBM-2	TEST	pos	-
0168ASP-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0175AMC-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0051BSC-ODIII-4	GLIOMA	AST-III	Aggressive-GLIOMA
0048BSC-ODIII-1	TEST	neg	-
0049BSC-ODIII-2	TEST	neg	-
0050BSC-ODIII-3	TEST	neg	-
0053BTN-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0054BTN-GBM-2	TEST	pos	-
0055BTN-GBM-3	TEST	pos	-
0058BTN-GBM-6	TEST	neg	-
0060BLA-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0059BLA-GBM-1	TEST	neg	-
0061BLA-GBM-3	TEST	neg	-
0062BLA-GBM-4	TEST	neg	-
0065BDP-GBM-3	GLIOMA	GBM	Aggressive-GLIOMA
0063BDP-GBM-1	TEST	neg	-
0064BDP-GBM-2	TEST	neg	-
0066BDP-GBM-4	TEST	pos	-
0067BDP-GBM-5	TEST	pos	-
0072BGD-ODIII-5	GLIOMA	OAST-III	Aggressive-GLIOMA
0068BGD-ODIII-1	TEST	pos	-
0069BGD-ODIII-2	TEST	pos	-
0070BGD-ODIII-3	TEST	pos	-
0073BGD-ODIII-6	TEST	pos	-
0071BGD-ODIII-4-0	TEST	neg	-
0074BAJ-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0075BAJ-GBM-2	TEST	pos	-
0076BAJ-GBM-3	TEST	pos	-
0077BAJ-GBM-4	TEST	pos	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0080BLM-GBM-3	GLIOMA	GBM	Aggressive-GLIOMA
0078BLM-GBM-1	TEST	pos	-
0079BLM-GBM-2	TEST	pos	-
0081BLM-GBM-4	TEST	pos	-
0082BLM-GBM-5	TEST	pos	-
0083BLM-GBM-6	TEST	pos	-
0086BLA-ODII_III-3	GLIOMA	ODG-II_III	Aggressive-GLIOMA
0085BLA-ODII_III-2	TEST	pos	-
0087BLA-ODII_III-4	TEST	neg	-
0088BLA-ODII_III-5	TEST	neg	-
0089BLA-ODII_III-6	TEST	neg	-
0084BLA-ODII_III-1-0	TEST	neg	-
0090BPG-ODII_III-1	GLIOMA	OASTII-III	Aggressive-GLIOMA
0091BPG-ODII_III-2	TEST	pos	-
0092BPG-ODII_III-3	TEST	pos	-
0093BPG-ODII_III-4	TEST	neg	-
0097BCA-GBM-4	GLIOMA	GBM	Aggressive-GLIOMA
0094BCA-GBM-1	TEST	pos	-
0095BCA-GBM-2	TEST	pos	-
0096BCA-GBM-3	TEST	pos	-
0098BGS-AST-1	GLIOMA	AST-II	Benign-GLIOMA
0099BGS-AST-2	TEST	pos	-
0100BGS-AST-3	TEST	pos	-
0101BGS-AST-4	TEST	pos	-
0103BEL-ODII-2	GLIOMA	OAST-II	Benign-GLIOMA
0102BEL-ODII-1	TEST	neg	-
0104BEL-ODII-3	TEST	pos	-
0105BPV-ODII-1	GLIOMA	ODG-II	Benign-GLIOMA
0106BPV-ODII-2	TEST	pos	-
0107BPV-ODII-3	TEST	pos	-
0108BPV-ODII-4	TEST	pos	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0113BLB-GBM-4	GLIOMA	GBM	Aggressive-GLIOMA
0111BLB-GBM-2	TEST	pos	-
0114BMM-AST-1	GLIOMA	AST-II	Benign-GLIOMA
0115BMM-AST-2	TEST	pos	-
0116BMM-AST-3	TEST	pos	-
0117BMM-AST-4	TEST	pos	-
0001BAL-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0003BBF-ODIII-2	GLIOMA	ODG-III	Aggressive-GLIOMA
0002BBF-ODIII-1-0	TEST	pos	-
0004BBM-ODIII-1	GLIOMA	ODIII_GBM	Aggressive-GLIOMA
0005BBM-ODIII-2	TEST	pos	-
0006BCG-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0007BEJ-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0009BEJ-GBM-3	TEST	pos	-
0010BEJ-GBM-4	TEST	pos	-
0011BEJ-GBM-5	TEST	pos	-
0013BEJ-GBM-7	TEST	pos	-
0014BEJ-GBM-8	TEST	pos	-
0015BEJ-GBM-9	TEST	pos	-
0016BEJ-GBM-10	TEST	pos	-
0018BHJ-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0017BHJ-GBM-1	TEST	pos	-
0019BLM-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0020BMO-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0021BMO-GBM-2	TEST	neg	-
0022BMO-GBM-3	TEST	neg	-
0023BMY-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0024BMJ-ODII_III-1	GLIOMA	ODG-III	Aggressive-GLIOMA
0025BMJ-ODII_III-2	TEST	pos	-
0026BNC-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0028BSJ-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0027BSJ-GBM-1	TEST	neg	-
0030BSS-GS-2	GLIOMA	GS	Aggressive-GLIOMA
0031BAB-ODIII-1	GLIOMA	ODG-III	Aggressive-GLIOMA
0032BAB-ODIII-2	TEST	pos	-
0035BDJ-ODII_III-3	GLIOMA	ODG-II_III	Aggressive-GLIOMA
0033BDJ-ODII_III-1	TEST	pos	-
0034BDJ-ODII_III-2	TEST	pos	-
0036BDJ-ODII_III-4	TEST	pos	-
0041BPD-GBM-5	GLIOMA	GBM	Aggressive-GLIOMA
0038BDP-GBM-2	TEST	neg	-
0039BDP-GBM-3	TEST	neg	-
0042BMR-ODIII-2	GLIOMA	ODG-III	Aggressive-GLIOMA
0043BMR-ODIII-3	TEST	pos	-
0044BWA-ODIII	GLIOMA	ODG-III	Aggressive-GLIOMA
0045BBL-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0046BBL-GBM-2	TEST	pos	-
0047BBL-GBM-3	TEST	pos	-
0048BBM-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0050BBR-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0049BBR-GBM-1	TEST	pos	-
0051BBR-GBM-3	TEST	pos	-
0052BBR-GBM	TEST	pos	-
0058BGJ-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0059BGJ-GBM-2	TEST	pos	-
0060BHB-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0061BHA-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0062BHA-GBM-2	TEST	pos	-
0064BHJ-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0063BHJ-GBM-1	TEST	pos	-
0065BHC-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0067BJJ-GBM-1	TEST	neg	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0068BJJ-GBM-2	TEST	neg	-
0069BJJ-GBM-3	TEST	neg	-
0070BLC-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0073BLM-GBM-4	GLIOMA	GBM	Aggressive-GLIOMA
0071BLM-GBM-1	TEST	pos	-
0072BLM-GBM-3	TEST	pos	-
0075BLM-GBM-2	TEST	neg	-
0076BLM-GBM-3	TEST	pos	-
0077BLM-GBM-4	TEST	pos	-
0079BLM-GBM-6	TEST	pos	-
0080BLM-GBM-7	TEST	pos	-
0081BMJ-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0082BMJ-GBM-2	TEST	pos	-
0084BLM-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0086BOL-GBM-2	GLIOMA	GBM	Aggressive-GLIOMA
0085BOL-GBM-1	TEST	pos	-
0087BOL-GBM-3	TEST	pos	-
0088BRC-GBM	GLIOMA	GBM	Aggressive-GLIOMA
0089BSA-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0090BSA-GBM-2	TEST	pos	-
0091BSA-GBM-3	TEST	pos	-
0093BSR-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0094BSR-GBM-2	TEST	pos	-
0095BSM-GBM-1	GLIOMA	GBM	Aggressive-GLIOMA
0096BSM-GBM-2	TEST	pos	-
0097BSM-GBM-3	TEST	pos	-
0099BSM-GBM-5	TEST	neg	-
0101BBG-ODIII-2	GLIOMA	ODG-III	Aggressive-GLIOMA
0103BNC-ODII-1	GLIOMA	ODG-II	Benign-GLIOMA
0104BNC-ODII-2	TEST	pos	-
0105BNC-ODII-3	TEST	pos	-

Table A.1 continued from previous page

Identifier	Group	Pathology	Pathologic classification
0106BNC-ODII-4	TEST	pos	-
0107BRU-ODII-1	GLIOMA	ODG-II	Benign-GLIOMA
0108BRU-ODII-2	TEST	pos	-
0053BGJ-AST-2-0	GLIOMA	GBM	Aggressive-GLIOMA
0054BGJ-AST-3-0	TEST	pos	-
0055BGJ-AST-4-0	TEST	pos	-
0056BGJ-AST-5-0	TEST	pos	-
0057BGJ-AST-6-0	TEST	pos	-
0005PT2-SCL-2	CONTROL	CTX	Control
0007PT3-MLF-1	CONTROL	CTX	Control